Abstract

A method for approximating an object's pose from a camera generated image of a scene is performed by first extracting a binary map from the image. The binary map is filtered to include silhouettes of objects located within a predetermined range of distances from the camera. An initial binary shape template may be applied to the binary map to locate potential target object silhouettes. Iterative stages of binary templates are applied to the each target object silhouette that represent a range of poses of the target object. Each stage of templates has higher spatial fidelity than the previous stage and poses corresponding to templates that do not sufficiently match the silhouette are eliminated from consideration. The target object's pose is approximated based on a set of templates that best matches the target object silhouette.